

《深度学习》

课程报告

班级：信研 2305

姓名：韩龙坤

学号：2023200806

特征级融合算法综述

摘 要

本文对特征级数据融合算法进行了分类概述，分布从基于概率统计的融合算法、基于逻辑推理的融合算法、基于特征抽取的融合算法、基于搜索的融合算法和基于神经网络的融合算法做了归纳，并且对数据融合的未来研究方向进行了总结论述。

关键词： 数据融合、算法、综述

目录

特征级融合算法综述

1 简介

数据融合已不是新鲜的概念，通过融合来自多个传感器的数据和相关信息，实现比单传感器更准确的判断。目前数据融合在军事和非军事领域都有着广泛的应用，在军事领域广泛应用于自动目标识别、自动驾驶导航、遥感、战场监测等，在非军事领域广泛应用于环境监视、机器人技术以及医疗技术。

其实人类本身就有着卓越的数据融合能力，通过视觉、味觉、触觉、气味，人类可以对食物的喜好做出准确的判断，而仅凭一种感觉，我们就无法对食物的喜好做出这样的判断。同样地，在军事上，希望在空对空防御和地对空防御中，综合地基和飞机电磁信号数据做出更准确的预警，在非军事领域，比如医疗领域，我们希望通过 X 射线、核磁共振、目视检查等多种数据进行综合处理来做出更准确的疾病诊断。

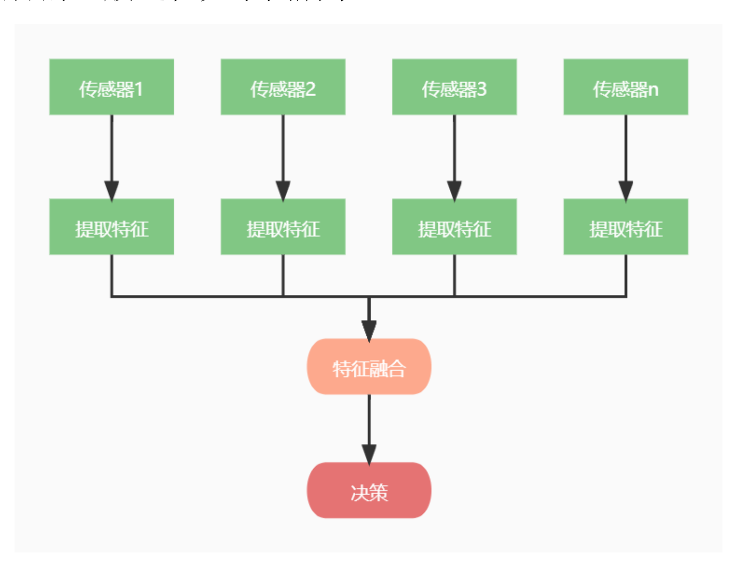
近年来，随着传感器技术的发展，信息表现形式多样化，信息数量巨大化，信息关系复杂化，以及要求信息处理的及时性、准确性和可靠性都是前所未有的。这就是使得利用计算机技术对获得的多源信息在一定的准则下进行分析、优化综合以完成所需的估计和决策（即多传感器信息融合技术）得以迅速发展。信息融合可以描述为：综合多源的信息，得到高品质的有用的信息。各种单一的传感器往往不能从场景中提取足够的信息，以至于很难甚至无法独立获得对一副场景的全面描述。需要多传感器的同时获取目标数据进行融合分析，才可有效的进行分类识别决策。多传感器信息融合又分为像素级融合、特征级融合、决策级融合三个层次。

特征级融合是从各个传感器获取的信息中提取目标特征，然后通过融合算法获取融合目标特征量，进行目标分类识别。特征级融合目标识别，既能保持足够数量的目标有效信息，去除冗余信息，又提高了目标识别的精确性。因此，从目标识别的实时性、计算量和识别精确性等方面考虑，本文以特征级融合研究作为重点。近几年特征级融合方法开始受到人们的重视，并且取得了许多成果。

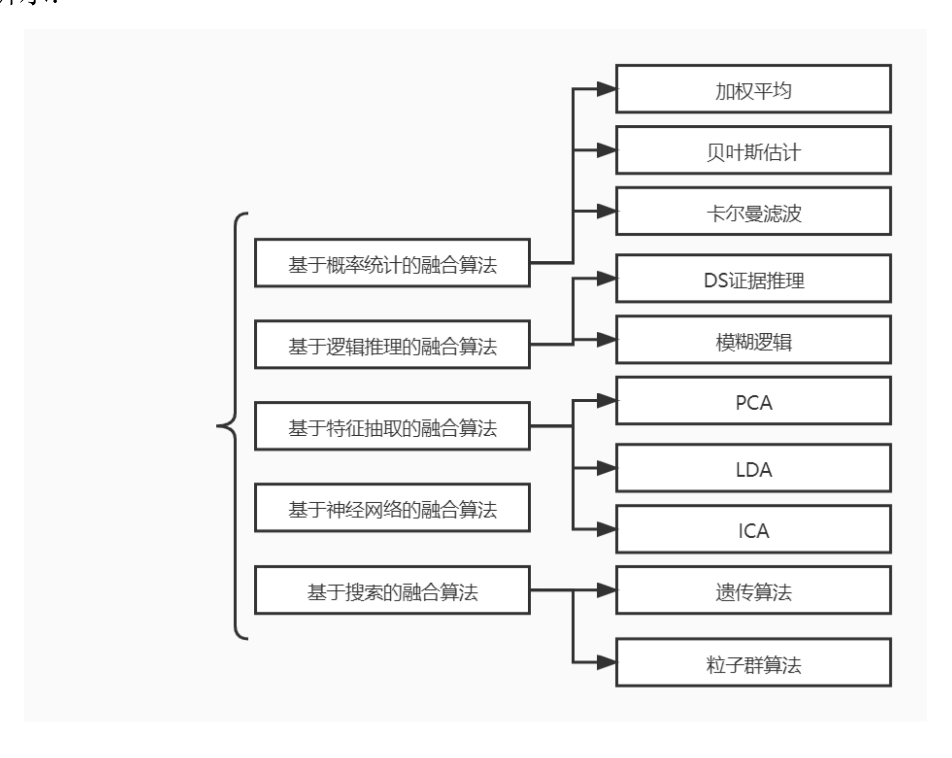
假设有 n 个传感器，第 i 个传感器的获取的特征为$\{d_1, d_1 \cdots d_n\}$（i 从 1 到 n，

$num(i)$ 表示第 i 个传感器提取特征个数），现在特征级融合目标就是要通过某种算法将 n 个传感器获取的所有特征融合为高品质的有用的信息 $\{d_1, d_1 \cdots d_m\}$ (其中 m 为最终融合得到的特征向量的特征个数)。然后将得到的融合特征作为进一步决策的依据。

特征级融合的一般过程如下图所示：



目前特征级融合方法大致分为五大类：概率论统计方法、逻辑推理方法、神经网络方法、基于特征抽取的融合方法和基于搜索的融合方法。分类细节如下图所示:

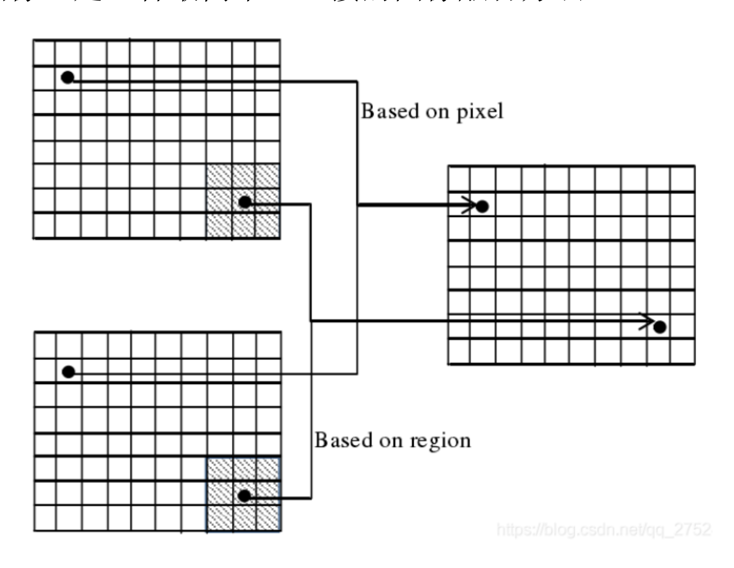


2 融合算法

2.1 基于概率统计的融合算法

2.1.1 加权平均

假设有 n 个传感器，用同一种方法进行特征提取后，对于包含 n 个维度的特征值 x 进行加权平均，用它作为该维度融合后的特征值。其中，权重可看成不同传感器准确性的度量。例如，最常见的加权平均是对图像的数据融合，对于三通道图像生成单通道图像是将源图像像素的三个通道灰度值进行线性加权平均，生成新的融合图像，是一种最简单、直接的图像融合方法。



该方法简单直观，但是必须事先对各个传感器进行详细的分析，获取它的权重，需要一定的先验知识。并且在不同特征维度上每个传感器的准确性都不一样，所以权重的获取成为主要难点。

2.1.2 贝叶斯估计

多贝叶斯估计法贝叶斯估计为数据融合提供了一种手段，是融合 静态环境中多传感器高层信息的常用方法。它使传感器信息依据 概率原则进行组合，测量不确定性以 条件概率表示，当传感器组的观测坐标一致时，可以直接对传感器的数据进行融合，但大多数情况下，传感器测量数据要以间接方式采用贝叶斯估计进行数据融合。多贝叶斯估计将每一个传感器作为一个贝叶斯估计，将各个单独物体的关联概率分布合成一个联合的后验的概率分布函数，通过使用联合分布函数的似然函数为最小，提供多传感器信息的最终融合值，融合信息与环境的一个先验模型提供整个环境的一个特征描述。贝叶斯估计方法在用于多传感器信息融

合时，是将多传感器提供的各种不确定信息表示为概率，并利用概率论中贝叶斯条件概率公式对他们进行处理的一种方法。

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

在先验概率已知的情况下，贝叶斯准则是最佳的融合准则，可给出精确融合结果。但是在实际应用中，各个传感器很难获得所需的先验概率，这一点大大限制贝叶斯准则的应用。

2.1.3 卡尔曼滤波

卡尔曼滤波主要用于融合低层次实时动态多传感器冗余数据。该方法用测量模型的统计特性递推。算法中涉及到滤波，也涉及到对下一时刻数据的预测。卡尔曼滤波由一系列递归数学公式描述。它提供了一种高效可计算的方法来估计过程的状态，并使估计均方误差最小。卡尔曼滤波器应用广泛且功能强大：它可以估计信号的过去和当前状态，甚至能估计将来的状态，即使并不知道模型的确切性质。

卡尔曼滤波也可以被认为是一种数据融合算法，已有 50 多年的历史，是当今使用最重要和最常见的数据融合算法之一。卡尔曼滤波的巨大成功归功于其小的计算需求，优雅的递归属性以及作为具有高斯误差统计的一维线性系统的最优估计器的状态。

卡尔曼滤波只能减小均值为 0 的测量噪声带来的影响。只要噪声期望为 0，那么不管方差多大，只要迭代次数足够多，那效果都很好。反之，噪声期望不为 0，那么估计值就还是与实际值有偏差。

2.2 基于逻辑推理的融合算法

逻辑推理已经被广泛地应用于多传感器融合中，和概率统计方法相比，逻辑推理的方法对信息的表示和处理更加接近于人类的思维方式，比较适合于高层次的融合。逻辑推理方法主要有：证据推理和模糊逻辑方法。

2.2.1 DS 证据推理

证据理论是由 Dempster 于 1967 年提出的，后由 Shafer 加以扩充和发展，所以证据理论又称为 DS 理论。证据理论可处理由“不知道”所引起的不确定性。它采用信任函数而不是概率作为度量，通过对一些事件的概率加以约束以建立信

任函数，而不必说明精确的 难以获得的概率，当约束限制为严格的概率时，它就成为概率论。

证据理论具有以下一些优点：(1)证据理论具有比较强的理论基础，既能处理随机性所导致的不确定性，又能处理模 糊性所导致的不确定性。(2)证据理论可以依靠证据的积累，不断地缩小假设集。(3)证据理论能将“不知道”和“不确定”区分开来。(4)证据理论可以不需要先验概率和条件概率密度。证据理论的主要缺点是：(1)证据理论具有潜在的指数复杂度。(2)在推理链较长时，使用证据理论很不方便。这是因为在应用证据理论时，必须首先 把相应于每个步骤和证据的信任函数变换成一个一般的识别框架，然后再应用 Dempster 组合规则，当推理步骤增加时，由于最后结果的信任函数的焦元结构的复杂性也相应增加，所 以 DempMer 规则的递归应用就会感到卜分困难。(3) Dempster 组合规则具有组合灵敏性,有时，基本概率赋值一个很小的变化都可能导 致结果很大的变化。此外，使用 Dempster 组合规则，要求证据是独立的，这个要求有时使用起来很不方便。

2.2.2 模糊逻辑

模糊逻辑是多值逻辑，通过指定一个 0 到 1 之间的实数表示真实度，相当于隐含算子的前提，允许将多个传感器信息融合过程中的不确定性直接表示在 推理过程中。如果采用某种系统化的方法对融合过程中的不确定性进行 推理建模，则可以产生一致性 模糊推理。与概率统计方法相比，逻辑推理存在许多优点，它在一定程度上克服了概率论所面临的问题，它对信息的表示和处理更加接近人类的思维方式，它一般比较适合于在高层次上的应用（如决策），但是，逻辑推理本身还不够成熟和系统化。此外，由于逻辑推理对信息的描述存在很大的主观因素，所以，信息的表示和处理缺乏客观性。模糊集合理论对于数据融合的实际价值在于它外延到 模糊逻辑，模糊逻辑是一种 多值逻辑，隶属度可视为一个数据真值的不精确表示。在 MSF 过程中，存在的不确定性可以直接用 模糊逻辑表示，然后，使用多值逻辑推理，根据模糊集合理论的各种演算对各种命题进行合并，进而实现数据融合。

2.3 基于特征抽取的融合算法

多变量问题是经常会遇到的。变量太多，无疑会增加分析问题的难度与复杂性。 在许多实际问题中，多个变量之间是具有一定的相关关系的。因此，能否

在各个变量之间相关关系研究的基础上，用较少的新变量代替原来较多的变量，而且使这些较少的新变量尽可能多地保留原来较多的变量所反映的信息？

因此，当输入的特征维度太大并且存在很大的冗余时，很难对其进行有效处理，为了对特征更好的分析，把当前特征空间投影到另外一个特征空间，从而达到降维的效果。主要有主成分分析法（PCA）、核 PCA、线性判别式分析（LDA）、独立成分分析（ICA）或者是改进后的典型相关分析（CCA）。

2.3.1 PCA

2.3.1.1 PCA 简介

PCA 由 Karl Pearson 于 1901 年发明，作为力学中主轴定理的类比。后来由 Harold Hotelling 在 20 世纪 30 年代独立开发并命名。

根据应用领域，它还被称为信号处理中的离散 Karhunen-Loève 变换（KLT），多变量质量控制中的 Hotelling 变换，机械工程中的正交分解（POD），X 的奇异值分解（SVD），线性代数中的 XTX 特征值分解（EVD），气象科学中的因子分析（关于 PCA 和因子分析之间差异的讨论），Eckart-Young 定理（Harman，1960）或经验正交函数（EOF），噪声和振动谱分解中的经验特征函数分解（Sirovich，1987），经验成分分析（Lorenz，1956），准谐波模式（Brooks et al，1988），以及结构动力学中的经验模态分析。

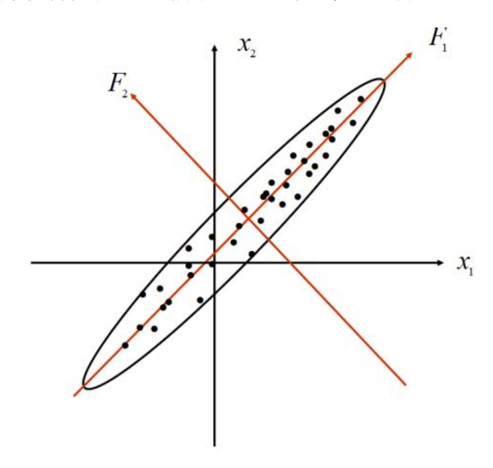
PCA 是一种降维的统计方法，它借助于一个正交变换，将其分量相关的原随机向量转化成其分量不相关的新随机向量。PCA 的原理是设法将原来变量重新组合成一组新的相互无关的几个综合变量，同时根据实际需要从中可以取出几个较少的总和变量尽可能多地反映原来变量的信息的统计方法，是数学上处理降维的一种方法。

2.3.1.2 PCA 几何解释

先假定数据只有二维，即只有两个变量，它们由横坐标和纵坐标所代表；因此每个观测值都有相应于这两个坐标轴的两个坐标值；

如果这些数据形成一个椭圆形状的点阵（这在变量的二维正态的假定下是可能的）。椭圆有一个长轴和一个短轴。在短轴方向上，数据变化很少；在极端的情况，短轴如果退化成一点，那只有在长轴的方向才能够解释这些点的变化了；这样，由二维到一维的降维就自然完成了。当坐标轴和椭圆的长短轴平行，那么

代表长轴的变量就描述了数据的主要变化，而代表短轴的变量就描述了数据的次要变化。但是，坐标轴通常并不和椭圆的长短轴平行。因此，需要寻找椭圆的长短轴，并进行变换，使得新变量和椭圆的长短轴平行。



如果长轴变量代表了数据包含的大部分信息，就用该变量代替原先的两个变量（舍去次要的一维），降维就完成了。对于多维变量的情况和二维类似，也有高维的椭球，只不过无法直观地看见罢了。首先把高维椭球的主轴找出来，再用代表大多数数据信息的最长的几个轴作为新变量；这样，主成分分析就基本完成了。

椭圆（球）的长短轴相差得越大，降维也越有道理。

2.3.1.3 PCA 算法步骤

表 2.1　PCA 算法步骤

设有 m 条 n 维数据
1 将原始数据按列组成 n 行 m 列矩阵 X
2 将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值
3 求出协方差矩阵 $C=\frac{1}{m}XX^T$
4 求出协方差矩阵的特征值及对应的特征向量
5 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前 k 行组成矩阵 P
6 $Y=PX$ 即为降维到 k 维后的数据

2.3.1.3 PCA 算法示例

原始数据为 2 维，有 5 个数据。组成的矩阵为：

$$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$

任务是将这个 2 维的数据集降到 1 维。首先按行均值化;

$$\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$

接下来去协方差矩阵：

$$C = \frac{1}{5}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$$

接下来取 C 的特征值和特征向量：

$$\lambda_1 = 2, \lambda_2 = \frac{2}{5}$$

$$c_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, c_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

c1 和 c2 可取任意非零实数。标准化后的特征向量为：

$$s_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, s_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

按特征值由大到小，将对应的特征向量按行由上往下写，得到矩阵：

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

由于我们要降到 1 维，即选取第一个主成分，即选取第一行，于是有：

$$Y = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

2.3.1.3 PCA 算法应用

主成分分析主要有以下几方面的应用：

（1）主成分分析能降低所研究的数据空间的维数。（2）多维数据的一种图

形表示方法。（3）由主成分分析法构造回归模型。即把各主成分作为新自变量代替原来自变量 x 做回归分析。（4）用主成分分析筛选回归变量。回归变量的选择有着重的实际意义，为了使模型本身易于做结构分析、控制和预报，好从原始变量所构成的子集合中选择最佳变量，构成最佳变量集合。用主成分分析筛选变量，可以用较少的计算量来选择量，获得选择最佳变量子集合的效果。

2.3.2 LDA

2.3.2.1 LDA 算法简介

给定带有标签的训练样本集，设法将样本投影到一条直线上，使得同类样本的投影点尽可能近，异类样本的投影点尽可能远。如果是进行分类，将新样本投影到同样的这条直线上，根据投影点的位置来确定新样本的类别。

将高维的模式样本投影到最佳鉴别矢量空间，以达到抽取分类信息和压缩特征空间维数的效果，投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离，即模式在该空间中有最佳的可分离性。

下面列出了执行线性判别分析的五个基本步骤。我们会在后面做更详细的讲解。

2.3.2.2 LDA 算法步骤

表 2.2　LDA 算法步骤

设有 m 条 n 维数据
1 计算数据集中不同类别数据的 d 维均值向量
2 计算散布矩阵，包括类间、类内散布矩阵。
3 计算散布矩阵的本征向量 $e_1, e_2 \cdots e_d$ 和对应的本征值 $\lambda_1, \lambda_2 \cdots \lambda_d$
4 将本征向量按本征值大小降序排列，然后选择前 k 个最大本征值对应的本征向量，组建一个 $d \times k$ 维矩阵——即每一列就是一个本征向量。
5 用这个 $d \times k$ 维本征向量矩阵将样本变换到新的子空间。这一步可以写作矩阵乘法 $Y = X \times W$ 。 X 是 $n \times d$ 维矩阵，表示 n 个样本；y 是变换到子空间后的 $n \times k$ 维样本。

2.3.3 ICA

2.3.3.1 理论背景

主成分分析来源于经典的鸡尾酒宴会问题。假设在 party 中有 n 个人，他们

可以同时说话，我们也在房间中一些角落里共放置了 n 个声音接收器用来记录声音。宴会过后，我们从 n 个麦克风中得到了一组数据，i 表示采样的时间顺序，也就是说共得到了 m 组采样，每一组采样都是 n 维的。我们的目标是单单从这 m 组采样数据中分辨出每个人说话的信号。

2.3.3.2 问题表述

设 X 是随机向量，且 $X \in R^{n\times 1}$，也就是说，X 有 n 个成员，每个成员是一个随机变量：

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

其中，x_i 是一个随机变量。

X 里的 n 个

将问题细化一下，有 n 个信号源，每一维都是一个人的声音信号，每个人发出的声音信号独立。A 是一个未知的混合矩阵，用来组合叠加信号 s。

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \mathrm{A} \begin{pmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{pmatrix}$$

其中，s_i 是一个随机变量，且凉凉相互独立，A 是一个满秩矩阵，且 $A \in R^{n\times n}$。

$$S = \begin{pmatrix} s_1 \\ s_2 \\ \cdots \\ s_n \end{pmatrix}$$

则：

$$X = AS$$

又有：

$$S = A^{-1}X$$

令：

$$\mathrm{W} = A^{-1}$$

则：

$$S = WX$$

其中， $W \in R^{n \times n}$ 。

记录随机向量 X 的值 m 次，则形成数据集：

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n1} & \dots & d_{nm} \end{pmatrix}$$

ICA 的目标 i 局势在只知道 D 的情况下，估算 A，W，S 的值。

2.4 基于搜索的融合算法

在当前特征空间的一些特征是强相关的，将所有的特征全部用于分类识别通常不是一种好的做法。基于搜索的方法旨在从源特征集中选择一个具有很强区分能力的特征子集。基于搜索的特征融合方法主要有基因遗传算法（GA）和粒子群算法（PSO）

2.4.1 遗传算法（GA）

2.4.1.1 算法描述

2.4.1.1.1 基于类内类间距离的可分性判据

要进行特征选择，首先要确定选择的准则，也就是如何评价选出的一组特征。确定了评价准则后，特征选择问题就变成从 D 个特征中选出使准则函数最优的 d 个特征（$d < D$）。

Fisher 线性判别采用了使样本投影到一维后类内离散度尽可能小，类间离散度尽可能大的准则来确定最佳的投影方向，这其实就是一个直观的类别可分性判据。可以用两类中的任意两两样本之间的距离的平均来代表两个类之间的距离，现在可以将其推导到多类情况。

令 $x_k^{(i)}, x_l^{(j)}$,分别为 ω_i 类及 ω_j 类中的 D 维特征向量，$\delta(x_k^{(i)}, x_l^{(j)})$ 为这两个向量间的距离，则各类特征向量之间的平均距离为：

$$J_d(x) = \frac{1}{2}\sum_{i=1}^{c} P_i \sum_{j=1}^{c} P_j \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \delta(x_k^{(i)}, x_l^{(j)})$$

式中 c 为类别数，n_i 为 ω_i 类中的样本数，n_j 为 ω_j 类中的样本数，P_i, P_j 是

相应类别的先验概率。

多维空间中两个向量之间有很多种距离度量，在欧式距离情况下有：

$$\delta(x_k^{(i)}, x_l^{(j)}) = (x_k^{(i)} - x_l^{(j)})(x_k^{(i)} - x_l^{(j)})^T$$

用 m_i 表示第 i 类样本集的均值向量:

$$m_i = \frac{1}{n_i}\sum_{k=1}^{n_i} x_k^{(i)}$$

用 m 表示所有各类的样本集的总平均向量:

$$m = \frac{1}{n_i}\sum_{i=1}^{c} P_i x_i$$

将上述式子代入到平均距离的公式得:

$$J_d(x) = \sum_{i=1}^{c} P_i \left[\frac{1}{n}\sum_{k=1}^{n_i}(x_k^{(i)} - m_i)(x_k^{(i)} - m_i)^T + (m - m_i)(m - m_i)^T \right]$$

也可以用下面定义的矩阵写出 $J_d(x)$ 的表达式,令:

$$\tilde{S}_b = \sum_{i=1}^{c} P_i (m - m_i)(m - m_i)^T$$

$$\tilde{S}_w = \sum_{i=1}^{c} P_i \sum_{k=1}^{n_i} (x_k^{(i)} - m_i)(x_k^{(i)} - m_i)^T$$

因为要保证类间离散度尽可能大，类内离散度尽可能小，故定义以下距离判据:

$$J_d(x) = \frac{tr(S_b)}{tr(S_w)}$$

基于距离的可分性判据定义直观、易于实现，因此比较常用。没采用基于类内类间距离的可分性判据，且根据公式可知， $J_d(x)$ 的值越大，表示特征的可分离性越好。

2.4.1.1.2 遗传算法

遗传算法把候选对象编码为一条染色体，在特征选择中，如 果目标是从 D 个特征中选择 d 个，则把所有特征描述为一条由 D 个 0/1 字符组成的字符串，0 代表该特征没有被选中，1 代表该特征被选中，这个字符串就叫做染色体，记作

m。显然，要求的是一条有且仅有 d 个的染色体，这样的染色体共有 C_D^d 种。 优化的目标被描述成适应度函数，每一条染色体对应一个适应度值 $f(m)$。可以用前面定义的基于类内类间距离的类别可分性判据 $J_d(x)$ 作为适应度。

2.4.1.2 算法步骤

2.4.1.2.1 初始化种群

假设数据集包含 60 个特征，一条染色体为一个 1*60 维的行向量，假设我们要从 60 个特征中选 30 个特征，则初始化的一个染色体为将一个零向量的随机 30 位变成 1，这样就从 60 维特征中随机选了 30 维，如下所示：

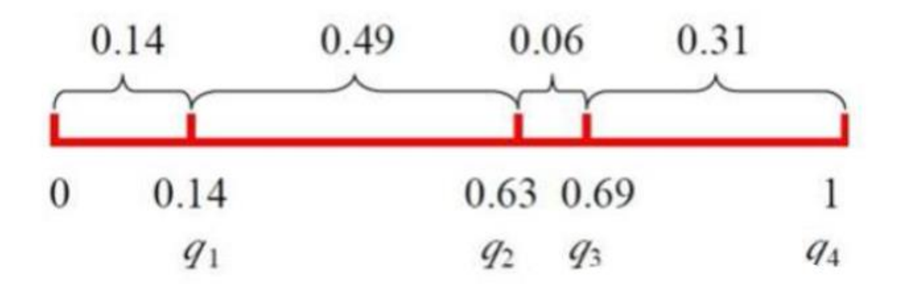
```
[ 1.  1.  0.  0.  1.  0.  0.  0.  0.  1.  1.  1.  0.  0.  0.  0.  1.  0.
  0.  0.  1.  1.  1.  0.  1.  1.  1.  1.  0.  0.  1.  0.  1.  0.  1.  0.
  0.  1.  1.  0.  1.  0.  1.  0.  1.  1.  0.  0.  0.  1.  1.  1.  0.  1.
  1.  0.  1.  0.  0.  1.]
```

重复上述过程 n 次，则可以得到一个有 n 条染色体的初代种群 $M(0)$，每条染色体都不尽相同。 经过 n 次计算之后，可以得到每条染色体的适应度值。

2.4.1.2.2 基于适应度值的选择

按照选择概率 $P(f(m))$ 对种群中的染色体进行采样，由采样出的染色体经过一定的操作繁殖出下一代染色体，组成下一代的种群 $M(t+1)$。

将种群中每条染色体的适应度值逐个累加，得到一些从 0 到 1 的区间，如图所示：



假设一个种群有 4 个条染色体，每条染色体的适应度值为 0.14、0.49、0.06、0.31，则将这些适应度值逐个累加起来，得到四个区间（0，0.14）、（0.14，0.63）、（0.63，0.69）、（ 0.69，1），每个区间的长度所代表着对应染色体的适应度值。我们从 0 到 1 中取一个随机数，该数落在哪个区间，就取哪条染色体。重复 n 次，

得到了基于上一代种群适应度值的新子代种群 $M(t+1)$，而且保证了种群的染色体数目不改变，恒为 n。

2.4.1.2.3 变异

基因突变是染色体的某一个位点上基因的改变。同样地，为了使变异之后染色体中特征数量不变，采用了以下方法：

先按一定的概率（称为变异概率）选择是否进行变异操作，若是，则随机从种群中选择一个个体，再随机地选择一个基因进行反转，若该基因由 1 变为了 0，则再随机选一个 0 变成 1，反之也执行同样的操作。直至遍历完种群中所有的个体。这样就能保证每个染色体的特征个数不会被改变。

2.4.1.2.4 重复迭代

在进行完选择、交叉、和变异操作之后，上一代的种群 $M(t)$ 已经变成了新一代的种群 $M(t+1)$。重复迭代，在遗传算法迭代的过程中，种群中的染色体会趋于所选特征数中的最优解，达到一定的迭代次数 t 后，算法停止，输出最终种群中适应度值最大的染色体，即完成了在 D 维特征中选择 d 个最优的特征。

2.4.2 粒子群算法（PSO）

2.4.2.1 算法描述

这是人们由鸟类觅食的集群活动的启发而提出来的一个启发式算法。首先，我们知道寻找一个合适的 0、1 串最简单也最全面的方法就是暴力枚举，也就是把所有的取值都计算一遍，这样是不现实的。共有 2^N 种可能，是一个 *NP-Hard* 问题。启发式算法就是在部分解中寻找一个最优解，当然这是一个局部最优解。但是，由于它只选择了部分解，因此时间空间开销都是较小的。

粒子群优化算法的思想是：假设空间中有一些粒子随机的分布在不同的空间位置上，每个粒子有两个属性——位置和速度。这两个属性的初始值都是随机的。每个粒子维护一个自身的最优点 *pbest*（最优指的是适应度函数取得最优时的位置值），整个粒子群维护一个全局的最优点 *gbest*（即所有粒子的 pbest 对应的适应度函数最大的那一个）。然后进行迭代，迭代的过程可以描述为粒子按照自身最优 *pbest* 的方向和全局最优 *gbest* 的方向按一定比例来改变速度，继而改变本次迭代粒子到达的位置。公式如下所示：

$$v_i^{t+1} = w \times v_i^t + c_1 \times rand \times (pbest - x_i^t) + c_2 \times rand \times (gbest - x_i^t)$$

速度公式可以这样理解：粒子在某次迭代时的速度由三个方面决定——自身原有的速度（上一次迭代得到的速度）、自身历史最佳点的速度、全局历史最佳点的速度。这三个分量按照一定的权重合成了本次迭代的速度。

位置公式：

$$x_i^{t+1} = x_i^t + v_i^{t+1}$$

位置公式就是简单的匀速直线运动公式，自定义一个迭代次数，按照这个方式进行迭代最终可能得到一个局部最优解，迭代结束时的全局最优解 gbest（适应度函数取最优时的位置值）就是算法得到的解。

在粒子群优化算法应用到特征选择的问题上，适应度函数可以定义为分类准确度的大小，也可以定义为分类错误率的大小。如果定义为前者，那么适应度函数取值最大时得到的是最优解；如果定义为后者，那么适应度函数取值最小时得到的是最优解，这里应用前文描述的基于类内类间距离的可分性判据作为适应度函数。

这里每个粒子的速度和位置都是 N 维向量，不要把思维局限在传统的物理学问题上认为速度和位置是一个值，这是 N 维空间的问题，每一维都有一个值。

2.4.2.2 二进制粒子群优化算法

根据前面的讲述，我们可以意识到——粒子群优化算法是一个解决连续空间问题的算法，而要想应用到特征选择问题上找到相要的 0、1 串，要做一些细微的修改使它能解决离散空间问题。当然，一种最简单的方法就是给位置值设定一个阈值，超过阈值就设为 1，否则就设为 0，这样就把连续空间的 N 维位置向量转换为离散空间的 0/1 串了。不过这样做缺乏物理解释，所以采用二进制粒子群优化算法。

它对粒子群优化算法的改进是：速度公式不变，增加了一个 *sigmoid* 函数，重新定义了位置公式。如下：

$$v_i^{t+1} = w \times v_i^t + c_1 \times rand \times (pbest - x_i^t) + c_2 \times rand \times (gbest - x_i^t)$$

$$x_i^{t+1} = \begin{cases} 1, rand < Sigmoid(v_i^{t+1}) \\ 0, otherwise \end{cases}$$

$$Sigmoid(v_i^{t+1})=\frac{1}{1+e^{-v_i^{t+1}}}$$

这样做的解释是，每次迭代有一定几率改变位置。

2.5 基于神经网络的融合算法

具有很强的容错性以及自学习、自组织及自适应能力，能够模拟复杂的非线性映射。神经网络的这些特性和强大的非线性处理能力，恰好满足了多传感器数据融合技术处理的要求。在多传感器系统中，各信息源所提供的环境信息都具有一定程度的不确定性,对这些不确定信息的融合过程实际上是一个不确定性推理过程。神经网络根据当前系统所接受的样本相似性确定分类标准，这种确定方法主要表现在网络的权值分布上，同时，可以采用神经网络特定的学习算法来获取知识，得到不确定性推理机制。利用神经网络的信号处理能力和自动推理功能，即实现了多传感器数据融合。

首先根据智能系统的要求以及传感器融合形式，选择神经网络模型和学习规则。同时将传感器的输入信息综合处理为一个总体输入函数，并将此函数映射定义为相关单元的映射函数，它通过神经网络与环境的交互作用把环境的统计规律反映到网络本身的结构中来；然后对传感器输出信息进行学习、理解、确定权值的分配，完成知识获取、信息融合，进而对输入模式做出解释，将输入数据向量转换成高层逻辑概念

（1）可以采用无监督学习，将所有的传感器得到的特征全部作为三层神经网络的输入，让输出层等于输入层，隐层节点即为所融合的特征。

（2）也可以采用监督学习方法，同样采用三层神经网，将所有的传感器得到的特征全部作为神经网络的输入，但是第三层作为决策层，通过带标签样本训练神经网络，隐层节点即为融合后特征。

神经网络的特点是：利用外部环境的特征信息可以实现知识的自动获取，能够将不确定性的复杂关系经过学习模拟出来，得到更高一层次的融合特征，由于神经网络具有并行大规模处理能力，使得系统信息处理加快。难点在于神经网络模型的建立，如隐层节点的个数如何确定等。

3 总结

传统的数据融合是指多传感器的数据在一定准则下加以自动分析、综合以完成所需的决策和评估而进行的信息处理过程。信息融合最早用于军事领域，定义为一个处理探测、互联、估计以及组合多源信息和数据的多层次多方面过程，以便获得准确的状态和身份估计、完整而及时的战场态势和威胁估计。它强调信息融合的三个核心方面：第一，信息融合是在几个层 次上完成对多源信息的处理过程，其中每一层次都表示不同级别的信息抽象；第二，信息融合包括探测、互联、相关、估计以及信息组合；第三，信息融合的结果包括较低层次上的状 态和身份估计，以及较高层次上的整个战术态势估计。

多传感器数据融合是人类或其他逻辑系统中常见的功能。人非常自然地运用这一能力把来自人体各个传感器（眼、耳、鼻、四肢）的信息（景物、 声音、气味、触觉）组合起来，并使用先验知识去估计、理解周围环境和正在发生的事件。

随着信息技术的发展和普及，特别是信息网络和信息高速公路的建设和应用，信息获取、综合分析和处理以及信息应用已经深入各行各业和社会的各个方面，为人们提供决策支持。由于信息化在各个方面展开，为了综合应用各种信息，需要对各方面的 信息技术进行围绕 因特网或信息高速公路的融合，以便产生新的增长点和开拓新的领域，不断地进行知识创新。

参考文献

[1] 吴小俊,曹奇英,陈保香等. 基于 Bayes 估计的多传感器数据融合方法研究[J]. 系统工程理论实践, 2000, 20(7): 45-48.

[2]Rachid Benmokhtar, Benoit Huet, Sid-Ahmed Berrani. Low-level feature fusion models for soccer scene classification. In Proceedings of the 2008 IEEE International Conference on Multimedia and Expo, ICME 2008, June 23-26 2008, Hannover, Germany. pages 1329-1332, IEEE, 2008. [doi]

[3]Cheng-Chi Wu, Wen-Li Lee, Yung-Chang Chen, Cheng-Hung Lai, Kai-Sheng Hsieh, Ultrasonic liver tissue characterization by feature fusion, Expert System s with Applications, Volume 39, Issue 10, August 2012, Pages 9389-9397, ISSN 0957-4174

[4]Mandeep Singh, Sukhwinder Singh, Savita Gupta, An information fusion bas ed method for liver classification using texture analysis of ultrasound images, Informa tion Fusion, Volume 19, September 2014, Pages 91-96, ISSN 1566-2535

[5]Pengfei Yu; Dan Xu; Hao Zhou, "Feature level fusion using palmprint and fi nger geometry based on Canonical Correlation Analysis," Advanced Computer Theor y and Engineering (ICACTE), 2010 3rd International Conference on , vol.5, no., pp.V 5-260,V5-264, 20-22 Aug. 2010 doi: 10.1109/ICACTE.2010.5579795

[6] Mangai Utthara ,Samanta Suranjana,Das Sukhendu 等. A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification [J].IETE Technical

[7]Cheng-Chi Wu; Wen-Li Lee; Yung-Chang Chen; Kai-Sheng Hsieh, "Evolution-Based Hierarchical Feature Fusion for Ultrasonic Liver Tissue Characterization," Bio medical and Health Informatics, IEEE Journal of , vol.17, no.5, pp.967,976, Sept. 201

[8] Cheng-Chi Wu, Wen-Li Lee, Yung-Chang Chen, Cheng-Hung Lai, Kai-Sheng Hsieh, Ultrasonic liver tissue characterization by feature fusion, Expert System s with Applications, Volume 39, Issue 10, August 2012, Pages 9389-9397, ISSN 0957-4174

[9] 孙虎,邱思棋,王鲁平等.特征级信息融合方法研究[J].电脑知识与技

术,2009,5(13):3342-3344.DOI:10.3969/j.issn.1009-3044.2009.13.006.

[10] 潘泉,于昕,程咏梅等.信息融合理论的基本方法与进展[J].自动化学报,2003,29(4):599-615.

[11]Bahador Khaleghi, Alaa Khamis, Fakhreddine O. Karray, Saiedeh N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information Fusion, Volume 14, Issue 1, January 2013, Pages 28-44, ISSN 1566-2535

[12]李先锋,朱伟兴,孔令东,花小朋. 基于 SVM 和 D-S 证据理论的多特征融合杂草识别方法[J]. 农业机械学报,2011,11:164-168+163.

[13]M. Dash, H. Liu, Feature selection for classification, Intelligent Data Analysis, Volume 1, Issues 1–4, 1997, Pages 131-156, ISSN 1088-467X

[14]任江涛, 孙婧昊, 黄焕宇,等. 一种基于信息增益及遗传算法的特征选择算法[J]. 计算机科学, 2006, 33(10):4.

[15]赵云, 刘惟一. 基于遗传算法的特征选择方法[J]. 计算机工程与应用, 2004, 40(15):3.

[16]任秀伟. 基于嵌入式稀疏特征选择策略的降维算法研究[D]. 湖北工业大学.

[17]. Mohajer, A., Yousefvand, M., Ghalenoo, E. N., Mirzaei, P., & Zamani, A. (2014). Novel approach to sub-graph selection over coded wireless networks with QoS constraints. IETE Journal of Research, 60(3), 203- 210.

[18]. Zhang, S., Madadkhani, M., Shafieezadeh, M., & Mirzaei, A. (2019). A novel approach to optimize power consumption in orchard WSN: Efficient opportunistic routing. Wireless Personal Communications, 108(3), 1611-1634.

[19]. Somarin, A. M., Alaei, Y., Tahernezhad, M. R., Mohajer, A., & Barari, M. (2015). An Efficient Routing Protocol for Discovering the Optimum Path in Mobile Ad Hoc Networks. Indian Journal of Science and Technology, 8(S8), 450-455.

[20]. Rezaei A, Dami S, Daneshjoo P. Multi-document extractive text summarization via deep learning approach. In2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI) 2019 (pp. 680- 685). IEEE.

[21]. Mohajer, A., Bavaghar, M., & Farrokhi, H. (2020). Reliability and mobility

load balancing in next generation self-organized networks: Using stochastic learning automata. Wireless Personal Communications, 114(3), 2389-2415.

[22]. Mohajer, A., Barari, M., & Zarrabi, H. (2016). Big Data-based Self Optimization Networking in Multi Carrier Mobile Networks. Bulletin de la Société Royale des Sciences de Liège, 85, 392-408.

[23]. S. &. V. R. Sindhuja, "A study on intrusion detection system of mobile ad-hoc networks," In Soft Computing for Problem Solving , pp. 307-316, 2020.

[24]. Dami S, Hatamchuri Z. Breast Cancer Prediction Using the Affinity Propagation Clustering with Regard to the Weights of Variables. Engineering Management and Soft Computing. 2018 Sep 23;4(2):27-39. [24]. Dami, S., & Yahaghizadeh, M. (2018, February). Efficient event prediction in an IOT environment based on LDA model and support vector machine. In 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) (pp. 135-138). IEEE.

[25]. Dami, S., Barforoush, A. A., & Shirazi, H. (2018). News events prediction using Markov logic networks. Journal of Information Science, 44(1), 91-109.. [26]. Dami, S. (2016). News Events Prediction Based on Casual Inference in First-Order Logic (FOL). Journal of Soft Comp

[26]. Young, C., Zambreno, J., Olufowobi, H., & Bloom, "Survey of automotive controller area network intrusion detection systems," IEEE Design & Test, vol. 36, pp. 48-55, 2019.

[27]. Javid, S., & Mirzaei, A. (2021). Presenting a Reliable Routing Approach in IoT Healthcare Using the Multiobjective-Based Multiagent Approach. Wireless Communications and Mobile Computing, 2021..

[28]. Somarin, A. M., Barari, M., & Zarrabi, H. (2018). Big data based self-optimization networking in next generation mobile networks. Wireless Personal Communications, 101(3), 1499-1518.

[29]. Mirzaei, A., & Najafi Souha, A. (2021). Towards optimal configuration in MEC Neural networks: deep learning-based optimal resource allocation. Wireless Personal Communications, 121(1), 221-243.

[30]. Rahimi, A. M., Ziaeddini, A., & Gonglee, S. (2021). A novel approach to

efficient resource allocation in loadbalanced cellular networks using hierarchical DRL. Journal of Ambient Intelligence and Humanized Computing, 1-15.